Remarks/Arguments:

Claims 1 and 8 have been amended. No new matter is introduced herein. Claims 1-15 are pending.

Claims 1 and 8 have been amended to clarify that the standard phonemic model is a group of phonemes. No new matter is introduced herein. Support for the amendment can be found, for example, at page 11, lines 2-6 of the original specification.

Claims 1-4, 6-11 and 13-15 have been rejected under 35 U.S.C. §103(a) as being unpatentable over Emori et al. (U.S. 6,934,681) in view of Chuang (U.S. 4,941,178). It is respectfully submitted, however, that these claims are patentable over the cited art for the reasons set forth below.

Claim 1, as amended, includes features neither disclosed nor suggested by the cited art, namely:

> ... for each of the frames, frequency-converting the respective acoustic feature parameter by filtering with a plurality of predetermined frequency conversion coefficients to form a corresponding plurality of frequency-converted feature parameters;
>
> determining, for each of the frames, a plurality of similarities or distances between each of the frequency-converted feature parameters and a standard phonemic model, the standard phonemic model being a group of phonemes;
>
> selecting at least one of the plurality of predetermined frequency conversion coefficients ... by using the determined plurality of similarities or distances for each of the frames; and
>
> normalizing the input utterance by frequency-converting the input utterance using the selected at least one predetermined frequency conversion coefficient. (Emphasis Added)

Claim 8 includes a similar recitation.

The features of claims 1 and 8 are described with reference to Figs. 8A and 8B of the subject specification. Applicants' claimed method of speaker normalization includes a step of frequency-converting, for each of the frames, the respective acoustic feature parameter of any input speech utterance with a plurality of frequency conversion coefficients. For example, in

Fig. 8, the plurality of frequency conversion coefficients are α_1-α_7. Applicants' claimed method also determines a plurality of similarities or distances between each of the frequency-converted feature parameters and each of the phonemes of a standard phonemic model, where the phonemic model includes a group of phonemes. For example, in Fig. 8A, in the first frame, coefficient α_4 (for phoneme /a/) has the highest similarity among coefficients α_1-α_7. Next, for each of the frames, one of the frequency conversion coefficients is selected, based on the plurality of similarities or distances. For example, for the first frame, coefficient α_4 is selected because corresponding phoneme /a/ is determined to be the maximum likelihood phoneme. The selected frequency conversion coefficient represents a frequency converting condition for normalizing the input utterance.

Emori et al. disclose, in Fig. 1, a spectrum converter including analyzer unit 1 for converting an input voice signal to an input pattern containing spectrum, elongation/contraction estimating unit 3 for outputting an elongation/contraction parameter and converter unit 2 for converting an input pattern using the elongation/contraction parameter (col. 7, lines 50-57). Elongation/contraction estimating unit 3 estimates the elongation/contraction parameter by using the spectrum in the input pattern (col. 7, lines 63-65). Estimating unit 3 obtains an alignment of the input pattern by using a hidden Markov model (HMM), and the elongation contraction parameter is calculated using the alignment, the HMM and the input pattern. (Col. 9, lines 9-22).

As described at col. 8, line 40 - col. 10, line 2, estimating unit 3 "executes elongation or contraction of the spectrum frequency without direct use of the spectrum" by executing a recursive conversion equation to estimate the elongation/contraction parameter. In other words, for every "predetermined interval of time," estimating unit 3 estimates one elongation/contraction parameter. Emori et al. teach that by estimating the elongation/contraction parameter, "it is not necessary to store various values in advance when determining the elongation/contraction parameter" and that it is not "necessary to execute distance calculation in connection with various values" (col. 7, line 63 - col. 8, line 2 and col. 10, lines 24-34).

As acknowledged by the Examiner, on page 2 of the Office Action, Emori et al. do not disclose or suggest, for each frame, frequency-converting an acoustic feature parameter by filtering with a plurality of predetermined frequency conversion coefficients, as required by claim 1 (emphasis added). Instead, Emori et al. teach estimating the elongation/contraction

parameter so that "repeat calculation as described before in the prior art is unnecessary, and analysis and other processes need to be executed only once" (col. 10, lines 29-34) (emphasis added). In fact, Emori et al. specifically teach that "it is not necessary to store various values in advance when determining the elongation/contract parameter" (col. 7, line 63 - col. 8, line 1).

Furthermore, Emori et al. do not disclose or suggest determining, for each of the frames, a plurality of similarities or distances between each of the frequency-converted feature parameters and a standard phonemic model, where the standard phonemic model is a group of phonemes, as required by claim 1 (emphasis added). Emori et al. are silent regarding this feature. Instead, Emori, et al. teach that estimating unit 3 uses an HMM corresponding to the voice signal inputted to the analyzer. Accordingly, the HMM may vary according to a change of the input voice signal. In contrast, according to Applicants' claimed method, any input utterance (even an unknown utterance) is processed with a plurality of frequency conversion coefficients and a standard phonemic model which includes a group of phonemes, such that the appropriate frequency conversion coefficients can be selected.

In addition, Emori et al. do not teach selecting at least one of the plurality of predetermined frequency conversion coefficients by using the determined plurality of similarities or distances for each of the frames, as required by claim 1 (emphasis added). Because Emori et al. estimate the elongation/contraction parameter, there is no need to select a coefficient by using a similarity of distance. In fact, Emori et al. specifically teach that it is not "necessary to execute distance calculation" (Col. 8, lines 1-2). Because Emori et al. do not disclose or suggest selecting at least one of a plurality of predetermined frequency conversion coefficients, Emori et al. cannot teach normalizing an input utterance using the selected predetermined frequency conversion coefficient, as required by claim 1. Thus, Emori et al. do not include all of the features of claim 1.

Chuang discloses, in Fig. 1A, a speech recognition system including slope filter estimate 16 and inverse filter 22 that provide a slope removal process to normalize the slope of LPC coefficients (Col. 4, lines 1-26 and Col. 6, line 63-Col. 7, line 52). The speech recognition system also includes all-pass filter 30, spectral warping 32 and time warping 34 for spectral normalization (after the slope normalization) and where the slope normalization and spectral normalization are regarded as speaker normalization. (Col. 8, lines 15-61 and Col. 9, lines 60-62). All-pass filter 30 provides expansion and compression of the LPC analysis results 24 along the frequency axis (Col. 8, lines 15-31 and Col. 8, lines 62-Col. 9, line 37).

Chuang does not make up for the deficiencies of Emori et al. because it does not disclose or suggest: 1) determining a plurality of similarities or distances between each of frequency-converted feature parameters and a standard phonemic model, where the standard phonemic model is a group of phonemes, 2) selecting at least one predetermined frequency conversion coefficient by using the determined similarities or distances for each of the frames or 3) normalizing the input utterance by frequency-converting the input utterance using the selected predetermined frequency conversion coefficient, as required by claim 1. Applicants have reviewed Col. 8, line 15-Col. 9, line 37 of Chuang, cited by the Examiner on pages 2-3 of the Office Action, and can find no disclosure of these features of claim 1. Applicants respectfully request that the Examiner either specifically point out where Chuang discloses these features or withdraw the rejection.

As described above, the combination of: 1) determining, for each of the frames, a plurality of similarities/distances between frequency-converted feature parameters and a standard phonemic model, where the standard phonemic model is a group of phonemes, 2) selecting at least one predetermined frequency conversion coefficient using the determined similarities/distances for each of the frames and 3) normalizing the input utterance by frequency-converting the input utterance using the selected predetermined frequency conversion coefficient is neither disclosed nor suggested by the cited art. Accordingly, allowance of claim 1 is respectfully requested.

Claims 2-4, 6 and 7 include all of the features of claim 1 from which they depend. Accordingly, claims 2-4, 6 and 7 are also patentable over the cited art.

Claim 8, although not identical to claim 1, includes features similar to claim 1 that are neither disclosed nor suggested by the cited art. Accordingly, allowance of claim 8 is respectfully requested for at least the same reasons as claim 1.

Claims 9-11 and 13-15 include all of the features of claim 8 from which they depend. Accordingly, claims 9-11 and 13-15 are also patentable over the cited art.

Applicants appreciate the indication, on page 5 of the Office Action, that claims 5 and 12 include allowable subject matter and would allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims. Applicants have not amended claims 5 and 12 at this time, however, because it is submitted that the base claims from which claims 5 and 12 respectively depend are allowable for the reasons set forth above.

In view of the amendments and arguments set forth above, the above-identified application is in condition for allowance, which action is respectfully requested.

Respectfully submitted,

Lawrence E. Ashery, Reg. No. 34,515
Attorney for Applicants

DMG/sh

Dated: October 7, 2009

P.O. Box 980
Valley Forge, PA 19482
(610) 407-0700

565247